

Original citation:

Webb, Colleen, Ferrari, Matthew, Lindstrom, Tom, Carpenter, Tim, Durr, Salome, Garner, Graeme, Jewell, Chris, Stevenson, Mark, Ward, Michael, Werkman, Marleen, Backer, Jantien and Tildesley, Michael J.. (2017) Ensemble modelling and structured decision-making to support emergency disease management. *Preventive Veterinary Medicine*, 138. pp. 124-133.

Permanent WRAP URL:

<http://wrap.warwick.ac.uk/84713>

Copyright and reuse:

The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions. Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Publisher's statement:

© 2017, Elsevier. Licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International <http://creativecommons.org/licenses/by-nc-nd/4.0/>

A note on versions:

The version presented here may differ from the published version or, version of record, if you wish to cite this item you are advised to consult the publisher's version. Please see the 'permanent WRAP url' above for details on accessing the published version and note that access may require a subscription.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

Manuscript Details

Manuscript number	PREVET_2016_61
Title	Ensemble Modelling and Structured Decision-making to Support Emergency Disease Management
Article type	Review Article

Abstract

Epidemiological models in animal health are commonly used as decision-support tools to understand the impact of various control actions on infection spread in susceptible populations. Different models contain different assumptions and parameterizations, and policy decisions might be improved by considering outputs from multiple models. However, a transparent decision-support framework to integrate outputs from multiple models is nascent in epidemiology. Ensemble modelling and structured decision-making integrate the outputs of multiple models, compare policy actions and support policy decision-making. We briefly review the epidemiological application of ensemble modelling and structured decision-making and illustrate the potential of these methods using foot and mouth disease (FMD) models. In case study one, we apply structured decision-making to compare five possible control actions across three FMD models and show which control actions and outbreak costs are robustly supported and which are impacted by model uncertainty. In case study two, we develop a methodology for weighting the outputs of different models and show how different weighting schemes may impact the choice of control action. Using these case studies, we broadly illustrate the potential of ensemble modelling and structured decision-making in epidemiology to provide better information for decision-making and outline necessary development of these methods for their further application.

Keywords	ensemble modelling, structured decision-making, policy, disease management, foot and mouth disease
-----------------	--

Corresponding Author	Colleen Webb
-----------------------------	--------------

Order of Authors	Colleen Webb, Matthew Ferrari, Tom Lindstrom, Tim Carpenter, Salome Durr, Graeme Garner, Chris Jewell, Mark Stevenson, Michael Ward, Marleen Werkman, Jantien Backer, Michael Tildesley
-------------------------	---

Abstract

Epidemiological models in animal health are commonly used as decision-support tools to understand the impact of various control actions on infection spread in susceptible populations. Different models contain different assumptions and parameterizations, and policy decisions might be improved by considering outputs from multiple models. However, a transparent decision-support framework to integrate outputs from multiple models is nascent in epidemiology. Ensemble modelling and structured decision-making integrate the outputs of multiple models, compare policy actions and support policy decision-making. We briefly review the epidemiological application of ensemble modelling and structured decision-making and illustrate the potential of these methods using foot and mouth disease (FMD) models. In case study one, we apply structured decision-making to compare five possible control actions across three FMD models and show which control actions and outbreak costs are robustly supported and which are impacted by model uncertainty. In case study two, we develop a methodology for weighting the outputs of different models and show how different weighting schemes may impact the choice of control action. Using these case studies, we broadly illustrate the potential of ensemble modelling and structured decision-making in epidemiology to provide better information for decision-making and outline necessary development of these methods for their further application.

Key Words: ensemble modelling, structured decision-making, policy, disease management, foot and mouth disease

Introduction

Transboundary livestock diseases can have devastating animal-health and economic impacts because such diseases are highly contagious, with the potential for rapid spread across geographic boundaries. Government agencies and livestock industries worldwide continue to develop and refine their policy and management actions in the face of such threats (e.g. Keeling et al., 2003; Schoenbaum and Disney, 2003; Tildesley et al., 2006; Willeberg et al., 2011; Yoon et al., 2006). Similar challenges exist more broadly in animal and human health, for example malaria (Murray et al., 2014), tuberculosis (Suen et al., 2014), and dengue fever (Wilder-Smith and Macary, 2014; Shaman et al., 2016). Decision-making when managing transboundary livestock diseases is complex; it must balance trade-offs amongst competing objectives, limited resources, and uncertainty in disease risk (Taylor, 2003). A variety of tools that incorporate data from empirical studies, previous outbreaks, and expert opinion are used to support science-based decision-making (Green and Medley, 2002; Woolhouse, 2003; Keeling, 2005), particularly for diseases such as foot and mouth disease (FMD) in non-endemic countries. Many tools used to understand the potential for infection spread and the effect of response actions on that spread inherently require an underlying predictive model of disease transmission (Kao, 2002; Woolhouse, 2003; Keeling, 2005; Garner and Hamilton, 2011; Mansley et al., 2011; Willeberg et al., 2011).

Given the complexity of disease ecosystems, it is difficult to describe all aspects of disease processes accurately within one model. Choices must be made regarding what to include and what to omit, how to implement specific processes, and how to parameterize them. Thus, model outputs upon which policy decisions are based differ owing to different modelling approaches, assumptions, and parameter estimates (Green and Medley, 2002). These model differences are

often justifiable. Different models may produce similar or quite different outputs that can all be considered plausible, where plausibility is often supported either from first principles and parameterization from known literature values in the absence of observed outbreak data or by the match between model outputs and the characteristics of observed outbreaks, when they are available. Variability among models is valuable because it captures uncertainty in the system and outbreak scenario, but reconciling variability can be difficult (Green and Medley, 2002; Keeling, 2005). Many fields, including weather forecasting, climate-change science, and medical science, use a diverse portfolio of models to indicate to decision-makers the amount of uncertainty in possible outcomes (Mangiameli et al., 2004; Palmer et al., 2004; Araujo and New, 2007). Thus, justified model diversity should be harnessed to produce cohesive policy recommendations from models, but this requires a method to incorporate potentially disparate outputs objectively from an ensemble of model outputs.

The idea of integrating model outputs to achieve a transparent decision-support framework has a relatively long history in weather forecasting (Sanders, 1963; Gneiting and Raftery, 2005 ;), hydrology (Cloke and Pappenberger, 2009; Velázquez et al., 2010), and climate-change modelling (Orsolini and Doblas-Reyes, 2003; Benestad, 2004; Palmer et al., 2004; Tebaldi and Knutti, 2007; Chandler, 2013). In medical sciences, multi-model approaches are used to assist physicians in making a medical diagnosis (Mangiameli et al., 2004; West et al., 2005). Examples of integrated approaches within the ecological literature are increasing (Niu et al., 2014) and include particle-filtering (Doucet et al., 2001) and Bayesian (Lindström et al., 2015) approaches to integrate multiple parameterizations of a single model; another approach is using integrated climate-change data to describe future environmental variables used as inputs into ecological models (Araujo and New, 2007; Barbet-Massin et al., 2009; Coetzee et al., 2009;

Thuiller et al., 2009; Maiorano et al., 2011). The latter approach has been applied in epidemiology where integrated climate projections were used to generate future environmental variables that drive predictions of disease incidence (Palmer et al., 2004; Thomson et al., 2006; Guis et al., 2012). To date, however, multiple model approaches have been applied only to a limited extent in public health (Thomson et al., 2006; Shaman et al., 2016) and in agriculture (Catelaube and Terres, 2005). Recent work suggests a way forward for multi-model, decision-support frameworks in epidemiology and animal health. This work focuses on ensemble modelling (Ward et al., 2007; Shaman and Karspeck, 2012; Lindström et al., 2015; Shaman et al., 2016) and structured decision-making (Shea et al., 2014; Probert et al., 2016), although available methods, at the time of writing, are at a preliminary stage.

Ensemble modelling (EM) combines model outputs to produce collectively a depiction of future states including uncertainty from several potential sources. Single-model ensembles use a single model structure but allow for different starting conditions and parameterizations whose outputs are combined to produce probability distributions of modelled outcomes (Tebaldi and Knutti, 2007). The mean of the probability distribution is the expected outcome, and credible intervals quantify uncertainty in the outcome. Two different single-model EM methods have been developed and applied in an epidemiological context to seasonal influenza (Shaman and Karspeck, 2012) and FMD (Lindström et al., 2015). Multi-model ensembles incorporate outputs from a set of structurally different models, referred to as an ensemble, that can incorporate different underlying processes and contribute to the uncertainty estimate (Tebaldi and Knutti, 2007). These methods are in development for epidemiology (e.g. Shaman et al., 2016), but we later present a preliminary case study addressing this methodological gap.

Structured decision-making (SDM) is a framework for analysing decisions by breaking them into component parts (Clemen, 1997). In doing so, the key impediments to making a decision are identified and effort can be focused on reducing uncertainty about relevant components. The goal is to identify the decision that mathematically maximizes (or minimizes) the specified objectives. By using a multi-model ensemble approach to SDM, uncertainty about underlying mechanisms and parameters may be incorporated in the decision process. SDM focuses on uncovering consensus as well as tradeoffs between underlying mechanisms/parameters (represented by different models) and choice of objectives. Hence, SDM is a method that uses the component parts of decision-making to organize or partition uncertainty across models and objectives into a format in which major sources of uncertainty can be identified and addressed. It has been used to facilitate decision-making in diverse fields such as organizational learning, the use and management of natural resources, adaptive management for pest control or biodiversity (Argyris and Schön, 1978; Hollings, 1978; Walters, 1986; Lee, 1993; Shea and Management, 1998; Parma, 1999; Shea et al., 2002; Williams et al., 2007; Williams, 2011; Keith et al., 2011; Williams et al., 2011) and recently in animal health (Probert et al., 2016).

Methodological development integrating EM and SDM is needed to create human- and animal-health decision-support frameworks that integrate multiple model results (Karemer et al., 2016; Lessler et al., 2016). A few studies have shown multiple model outputs side-by-side (Murray et al., 2012; Smith et al., 2012; Probert et al., 2016) or have truly integrated outputs from multiple parameterizations of a single model (Shaman and Karspeck, 2012; Lindström et al., 2015). However, these approaches are not well-established and methods are lacking to deal

with integration of multiple, policy-informative simulation models with complex model structure.

Our goal in this paper is to illustrate the potential of a combined multi-model EM and SDM approach and encourage further work in this area. We present two illustrative case studies; one highlighting the implementation of multi-model EM for an SDM scenario using a mock FMD outbreak simulated in Cumbria, UK, and one focusing on how to incorporate models with varying levels and types of plausibility into ensemble results by weighting the contribution of different models in an objective fashion using a mock FMD outbreak simulated in The Midlands and Wales, UK. We use an ensemble of FMD models that have been developed by a number of FMD-free countries that are engaged in preparedness planning (Ferguson et al., 2001; Keeling et al., 2001; Morris et al., 2001; Garner and Beckett, 2005; Harvey et al., 2007; Stevenson et al., 2013) because of the large economic losses associated with previous outbreaks. We first briefly describe the situation with FMD modelling. We then apply EM and SDM approaches to illustrate how they can be used to integrate the outputs from multiple models and inform policy and outbreak management in the two case studies. However, we stress that our goal is not to provide specific recommendations with respect to FMD and that our results should not be taken as a broad policy recommendation. Instead our goal is to illustrate how EM and SDM approaches could be more broadly applicable to both human- and animal-disease preparedness planning and response. We focus on FMD models because this is where our expertise lies and because it is an important transboundary livestock disease with appropriate existing model results that were available to us. In conclusion, we discuss the logistics of a fuller integration of EM and SDM and the potential benefits to disease response and preparedness planning.

Foot and mouth disease models

We focus here on stochastic, spatially-explicit simulations of FMD, which comprise the majority of models used to inform FMD policy in the last decade, e.g. AusSpread (Garner and Beckett, 2005; Beckett and Garner, 2007), the Central Veterinary Institute model (CVI, Backer et al., 2012), Exodis FMD (DEFRA, 2005), InterSpread Plus (Morris et al., 2001; Stevenson et al., 2013), the North American Animal Disease Spread Model (NAADSM, Harvey et al., 2007), and the Warwick model (Keeling et al., 2001; Tildesley et al., 2006). While each of these models simulates the spread of disease between geographical locations where groups of animals are managed as a single unit (i.e. farms), they differ in the way infection and disease transmission is implemented. Many of these models incorporate multiple, specific pathways of transmission and are generally designed to reflect the environment, production and marketing systems of the source country for the model. Transmission pathways of infectious diseases mostly depend on the biology of the disease and are similar within different countries. However, these models also have built in flexibility that means they can be reparameterized or restructured and thus many of them can and have been used for other countries or diseases. Examples of transmission mechanisms include livestock shipments, feed truck deliveries, wind borne movement and fence line contact. These models are often parameterized from empirical data collected during the course of FMD outbreaks in other countries, survey data and expert opinion. Models of this type include AusSpread, InterSpread Plus, and NAADSM. Other livestock disease models, such as CVI and Warwick, use phenomenological spatial kernels to represent a convolution of specific transmission pathways where the spatial kernel describes the neighbourhood of influence of an infectious location and the risk of disease transmission generally decreases as a function of distance from the focus of infection. The risk of infection is therefore based upon the location,

size and species composition of each premises as well as the distance between them. The parameters of the spatial kernel can be estimated based upon historical data (Keeling et al., 2001, Hayama et al., 2013). Exodis-FMD uses a mixture of spatial kernels and specific transmission pathways. In the interest of brevity, we do not describe further details of the models, but present a summary (Table 1) and rely on this summary, their policy relevance and peer-reviewed status as sufficient justification of the models since the work proposed here does not depend directly on the exact details of the models.

Within the context of FMD (and we suspect for other disease systems as well) the lack of a decision-support framework for integrating model outputs means that often a single model is used by analysts and policy makers or when multiple models are used their integration is informal. Although these informal integrations are generally regarded as appropriate, decision-making could be improved by more formal methods and transparency in how multiple model outputs are combined through EM and SDM.

The first steps of a multi-model approach were begun as part of the “QUADS” series of comparison studies (Dubé et al., 2006; Roche et al., 2014, Roche et al., 2015) in which results were compared for standardized scenarios across a suite of FMD models (AusSpread, CVI, Exodis FMD, InterSpread Plus, and NAADSM). The QUADS studies found that model results were similar across many--but not all-- of the scenarios considered; the QUADS studies also improved the understanding of individual models by highlighting the importance of model assumptions that generated outputs that differed from the rest of the model suite. This type of comparison was critical because it provides a logical starting point for fuller integration of outputs, e.g. EM and SDM. To illustrate EM and SDM, we focus on the models used in the QUADS studies plus one additional model (Warwick).

Case study one: Structured decision-making

Uncertainty in model outputs given a particular control action is sometimes of more interest than the predicted number of infected locations or epidemic duration (Yoon et al., 2006). The ensemble of model outputs encapsulates this uncertainty about the spatiotemporal dynamics of infection spread, which may be a limiting step in the decision process. SDM assists decision-making by incorporating this uncertainty while mathematically determining optimal management decisions given specified objectives (Shea et al., 2014). The first step in an SDM approach is to formalize the objectives, i.e. the fundamental goals that managers are trying to achieve through their actions. The objectives, e.g. minimizing loss of livestock, minimizing epidemic duration, minimizing economic costs, then provide a common measure by which to evaluate control actions implemented in each model in the ensemble.

For relatively simple decision-analysis problems, the objectives can be evaluated by generating a simulation experiment to project the outcome of all possible combinations of control actions and models under consideration. Because our goal is to provide a perspective on the use of SDM in epidemiology, we direct readers interested in more detailed methods to Probert et al. (2016). In this case study, we focus on three FMD models where the needed outputs were available to us: AusSpread, NAADSM, and Warwick (Table 2). Within the case studies, we anonymize model names because our focus is on ensemble methods and not model comparison. We illustrate SDM with a simple simulation experiment for a landscape consistent with Cumbria, UK (details in Appendix A) that determines the mathematically optimal decision for a given objective among five possible control actions in response to an FMD outbreak: 1) culling of infected premises (IPs) only; 2) culling of IPs and those that have been identified as at

206 risk because they have had contact with IPs (contact tracing); 3) culling of all farms within 3 km
207 of IPs in addition to IP culling; 4) vaccination of all farms within 3 km of IPs in addition to IP
208 culling; and 5) vaccination of all farms within 10 km of IPs in addition to IP culling. The model
209 outputs depend strongly on multiple factors specific to the scenario investigated here, such as
210 underlying farm demography, the level of efficiency in the implementation of control strategies
211 and constraints on control resources. Hence, policy recommendations from the case study are
212 specific to this scenario.

213 The output of each simulation was summarized with respect to three measures of the
214 outbreak: 1) the economic cost (see description in Appendix A) in terms of the re-imbursement
215 payments to producers for culled animals only, assuming that vaccinated animals are not
216 subsequently culled owing to vaccination (vaccinate-to-live); 2) the economic cost in terms of
217 the re-imbursement payments to producers for culled and vaccinated animals (i.e. assuming that
218 vaccinated animals will also be subsequently culled owing to vaccination); and 3) the duration of
219 the epidemic from the first detected case to the last animal culled or vaccinated, which would
220 reflect the economic costs associated with the disruption of trade due to export bans. Particularly
221 with respect to the vaccinate-to-live strategies, we highlight that these strategies have a number
222 of other impacts (e.g. on animal movement, trading bans and animal welfare) that are not
223 specifically captured in the outbreak measures used. The outcome of each control action was
224 simulated within the three models and the optimal action was taken as that which minimized the
225 outbreak duration (Table 2) or economic cost (Table 3). See Appendix A for details of the
226 simulations.

227 Here, all three FMD models predict the lowest mean cost due to livestock culled if a 10-
228 km ring vaccination action was applied – thus, although each model predicts different numbers

of cattle culled (Figure 1), the decision that minimizes that outcome is robust to model uncertainty. In contrast, if the objective was to minimize the duration of the outbreak – i.e. because of the larger economic costs of trade restrictions – the three models in the ensemble made differing predictions of the best control action: both models 1 and 2 recommended a 3-km culling ring, whereas model 3 recommended a 10-km vaccination ring (Table 2). This highlights that the important distinction is whether the transmission dynamics are more likely to behave like those of models 1 and 2 or like model 3, but distinguishing between models 1 and 2 would not affect the decision about the action to take. In the absence of empirical evidence supporting one model over another, policy-makers might set the initial policy as that which minimizes the expected objective with respect to model uncertainty; here, 3-km ring culling is the preferred option if the three models are given equal weight. If there is support for unequal weighting of projection models, this can easily be incorporated into the proposed framework by taking a weighted average of projected outcomes (i.e. an expectation relative to a probability model with unequal weights on projection models) (McDonald-Madden et al., 2010; Shea et al., 2014). There are many ways to arrive at unequal weights for projection models, ranging from goodness-of-fit to historical or contemporary surveillance data to expert opinion (McDonald-Madden et al., 2010; Shea et al., 2014). We present a novel approach to assessing model weights below.

Model uncertainty need not be the only factor limiting decision-making (Probert et al., 2016). The mathematically optimal decision is a consequence of interactions between the underlying model dynamics and the management objective. Table 3 illustrates the dependency of the least costly control action, with outcomes averaged over the three FMD models, for two different management objectives (i.e. measures of epidemic outcome). Clearly, when vaccination has a low cost (i.e. compensation is only required for infected and not for vaccinated

animals – vaccinate-to-live) an aggressive vaccination approach is favoured in all models. However, if producers must be compensated for vaccinated animals (vaccinate-to-die), then limited culling minimizes costs. Vaccination may incur additional costs not considered here, such as longer trade bans (Paarlberg et al., 2008; Anonymous, 2014) and, as seen above, more aggressive ring culling results in the shortest outbreaks, when averaged across all models (Table 2). Thus, by taking an ensemble approach, we can highlight consensus recommendations and the sensitivity of model output to the formulation of objectives that might have been confounded with model choice in a single model analysis (Probert et al., 2016). Total economic costs are arguably a more complete, and perhaps preferable, objective. However, their calculation requires a sophisticated economic analysis taking into account decisions made by trading partners that may itself have significant uncertainty. The specification of a full economic model for outbreak costs is beyond the scope of the current analysis, but we address the dependence of the analysis on alternative objectives in the General Discussion.

Case study two: Model weighting

In case study one, the contribution of each model was equally weighted and its influence spread uniformly (see also Murray et al., 2012; Smith et al., 2012). Here, we illustrate the application of the Bayesian Reliability Ensemble Average (BREA) method (Tebaldi et al., 2005) to epidemiology, which can take into account multiple influences on model weights (see Appendix B and Lindström et al., 2015 for technical details). The original BREA method estimates model weights based on agreement with observed data (bias criterion) and consensus between models (convergence criterion), which down-weights outliers. In the original climate change application of BREA (Tebaldi et al., 2005), the main quantity of interest was the

estimated current and future mean temperature. The framework was set up to allow for correlation between current and future temperature estimates, so that, for example, a model that under-predicts current mean temperatures might also do so for future mean temperatures. The BREA climate change example is analogous to the epidemiological problem where instead of current and future mean temperatures we substitute an outbreak quantity under the implemented control strategy and an alternative control strategy that a policy maker would like to compare (Lindström et al., 2015). This approach is easily expandable to consider multiple outbreaks and multiple, alternative control actions in epidemiological applications.

A major advantage is that BREA produces easily interpretable probability distributions for outbreak quantities (e.g., size, duration, economic costs) under two or more different control actions. The BREA framework promotes straightforward communication of uncertainty in outcomes and the effect of control actions rather than just the most likely outcome (Wade, 2000) or an equally-weighted, average outcome (as in Case Study 1). The BREA method is also technically appealing because it can be used for applications where relatively small amounts of data are available and model fitting-to-data is not required (Lindström et al., 2015). The weightings in the BREA method can be based on summary statistics (e.g. number of infected premises, outbreak duration, economic costs), which allows integration of models for which outputs are not necessarily of the same format (e.g. temporal or spatial scale). Thus, we anticipate that the BREA method will be broadly applicable in veterinary epidemiology.

Our case study incorporated simulations from a QUADS scenario outbreak consistent with the Midlands counties and Wales in the UK performed with five models: NAADSM, AusSpread, CVI, Exodis FMD, InterSpread Plus, and we further added the Warwick model to the ensemble. We used outbreak duration as the quantity of interest and focused on comparison of two control

actions from the QUADS studies (Roche et al., 2014; 2015): IP culling (scenario S0 in the QUADS studies: stamping out) and IP culling plus suppressive, prospective vaccination within one km around IPs (scenario V6 in the QUADS studies). See Appendix B for more details on the simulations. The original QUADS studies were based on standardized scenarios for model comparison as opposed to actual outbreak data. Thus, we were unable to implement the bias criterion aspect of estimated weights for this case study. Instead, we focus on comparison between equal-weighting as in Case Study 1 and weighting using the convergence criterion to down-weight outliers. We discuss the role of the bias criterion in estimating weights in the General Discussion below.

Figure 2 shows the mean individual-model outputs as well as the marginal posterior probabilities (probability distributions) of outbreak duration under the two considered weighting schemes: equal-weighting and weighting based on the convergence criterion (see Appendix B and Lindström et al., 2015 for technical details). Depending on the weighting scheme, the expected outbreak duration (posterior mean and 95% central credibility interval) is reduced by 44.5 [-4.2, 104.3] or 32.8 [0.2, 88.2] days when vaccination is implemented with equal-weighting and convergence-weighting respectively. When implementing the convergence criterion for weighting, the distributions are shifted towards the centre of the ensemble compared to equal-weighting. This formally down-weights outliers, providing a more conservative estimate of the reduction in duration with vaccination, which here indicates a positive effect of vaccination in the Midlands counties and Wales scenario. However, the probability distributions corresponding to either weighting scheme are wide, with estimated reduction ranging from little (or no) effect to several months. This stems from the discrepancy among the model predictions, and demonstrates the hazard of relying on a single model to inform policy.

As the number of outbreaks and control actions considered increases, the complexity of estimating convergence-weighting increases and would be extremely difficult to justify without a BREA-like approach. Returning to an issue raised in the Introduction, the assumption in this case study is that the weighting of models differs based on their similarity with other models. Models with lower weights in this context are not eliminated from the ensemble (instead they are down-weighted); and incorporating some influence of these models on the integrated predictions is justified given that their similarity (convergence) with other models in this case study differs under different control actions (e.g. in Figure 2 the green and cyan models are outliers under different control scenarios). Similarly if we had been able to include the bias-weighting in this case study, models would be further weighted with respect to their predictions of observed outbreak statistics (see Lindstrom et al., 2015 for a single-model example with both bias- and convergence-weighting).

General Discussion

Given the differences among modelling approaches, they sometimes appear to be in competition with one another (Kao, 2002; Woolhouse, 2003; Keeling, 2005; Garner and Hamilton, 2011). We suspect this competition largely comes from limited funding and constraints on how much model uncertainty can currently be incorporated into policy recommendations so that often a single model informs policy. However, model differences can be important characterizations of different risks in an outbreak, and uncertainty in these risks should be propagated to the evaluation of alternative actions. There is also growing interest in collaboration among different modelling teams (Dubé et al., 2007; Gloster et. al., 2010; Sanson et al., 2011) that serves to enhance emergency preparedness and builds confidence in model results. Ensemble approaches provide a way to use models representing different assumptions in

a complementary framework, thus emphasizing the potential for models to be mutually informative while propagating uncertainty in epidemic processes to the evaluation of actions.

Case Study 1 using SDM, and Case Study 2 using BREa produce qualitatively similar results: that the addition of ring vaccination with a relatively smaller radius results in shorter outbreaks (~30 days shorter) in expectation; but, the BREa analysis highlights that strong variation in outcomes within and between model projections results in very weak evidence that this intervention will differ from simple IP culling. However, our goal is not to recommend particular control actions for FMD, but to illustrate how control recommendations can be integrated across multiple models and objectives. Model predictions of the effectiveness of control will be highly dependent upon logistical capacities and it is therefore important to stress that the control strategies predicted to be optimal in this analysis according to the SDM approach may change as culling and vaccination capacities are varied. This phenomenon has been investigated in detail elsewhere for the Warwick model (Tildesley et al. 2006).

SDM, as illustrated in Case Study 1, focuses on the issues associated with the choice of objective and the potential for tradeoffs when multiple objectives are considered. One obvious choice of objective is total economic costs, as is reducing the risk of adverse events (Gerber et al., 2007). In the 2001 UK FMD outbreak, implementation of specific control actions was influenced by several factors throughout the epidemic, including the availability of resources, the perceived likelihood of spread and public perception of the impact of interventions (Andersen 2002). Hence, objectives associated with animal welfare (e.g. number of animals impacted), maintenance of culturally important lifestyles (e.g. number of family farms impacted), environmental damage (e.g. arising from the burial or burning of carcasses) and crisis fatigue (e.g. duration of the control period) may better reflect the objectives of the many stakeholders in

this decision. Exact specification of these objectives may only be possible with retrospective analysis in which data on direct outbreak costs as well as trade and additional other impacts are available. In response situations and for more open-ended preparedness planning scenarios, information on costs not directly associated with control actions can be difficult to specify. In these situations, direct measurements of the outbreak such as the number of animals infected, the number of premises infected and outbreak duration along with associated costs of these actions may be all that is available. Thus, there are multiple objectives that may be desirable to consider and understanding how tradeoffs among them interact with model uncertainty is the goal of SDM and of benefit in decision-making.

In contrast to SDM, BREA focuses on how to integrate multiple weighting schemes. Bias-weighting has been used for several single-model ensembles (Murray et al., 2012; Shaman and Karspeck, 2012; Lindström et al., 2015), and the next steps are to implement these methodologies for the type of multi-model ensembles illustrated in Case Study 2. Bias-weighting, based on the match of model predictions to observed data, is clearly an important way to incorporate the plausibility of models into an integrated policy recommendation. However, it should not be the sole consideration in all circumstances. Our experience is that models often perform differently in different situations, and there is no single best model in terms of prediction accuracy in all settings. Thus when considering alternative future control actions, i.e. for which observed data are unavailable, weighting based on bias relative to past observations alone may unnecessarily down-weight models that are more plausible for alternative control actions. Convergence-weighting, based on the match of model predictions to each other, is a complementary approach. The assumption here is that models that incorporate appropriate mechanisms, for example because they are based on established first principles, should behave

390 similarly. The incorporation of both bias- and convergence-weighting captures the tradeoff
391 between bias and precision in ensemble forecasts or predictions and would be our recommended
392 approach. Because BREA methods are Bayesian, expert opinion in the form of priors can also
393 be included (Kuhnert et al., 2010).

394 While EM and SDM methods individually facilitate the incorporation of multiple models
395 into decision-making, we advocate the development of methodologies that combine both
396 approaches by combining multiple objectives and weighting schemes. This is feasible within the
397 BREA framework and methods development is underway to expand the BREA framework with
398 bias- and convergence-weighting to multiple summary statistics. Multiple summary statistics are
399 often correlated, and this must be appropriately taken into account. However, different summary
400 statistics have different information content if not fully correlated. Thus, using a combination of
401 summary statistics will further improve predictions (as more information can be used) while
402 more fully incorporating tradeoffs among objectives and multiple weighting schemes. This
403 overall framework is highly flexible and can be applied in both preparedness and response
404 settings with potential expansion to address questions beyond alternative controls. Analogous
405 with climate change in which the goal is to capture current and future climate characteristics,
406 BREA could use current outbreak data to predict future outbreak characteristics, such as final
407 size and duration for proposed response scenarios. Further, this overall framework can be
408 extended to allow for adaptive decision-making; i.e. as with model weights in EM, real-time
409 observation may result in increased support for a subset of models within the ensemble and thus
410 decisions might be made with greater weight on the outputs of that subset (Williams et al., 2007;
411 Williams, 2011; Williams et al., 2011). As a given outbreak progresses, observations may
412 increasingly support the predictions of one model over the others, setting the stage for an

adaptive management approach (Williams et al., 2007; Williams et al., 2011; Williams, 2011; Shea et al., 2014) that shifts from the initial action that is robust to model uncertainty, to an action that is conditionally optimal for the best supported model.

There are many potential benefits to a combined EM and SDM approach simply in terms of the integration across models and objectives for more straightforward policy recommendations. Additionally, ensemble methods have improved prediction over single models in other areas of science (Palmer, et al., 2004; Gneiting and Raftery, 2005; Velazquez et al., 2010; Niu, et al. 2014). Our experience has been that the primary hurdles to integrating multiple models are not technical but logistical. Choice of plausible models to include in the ensemble is key as an ensemble of poor models can only produce poor predictions. The individual models are complicated, so organizing collaboration among modeling groups or training individuals to work across multiple models is both critical and challenging. For many transboundary animal diseases, including FMD, the data are international and confidential in nature and often government owned. Thus, negotiating international access and agreements for data sharing with modeling groups is also a challenge. A final challenge is developing an appropriate pipeline that works across different models for implementing standardized scenarios and standardized outputs of individual models for use in the ensemble model. We find that a formal feedback stage including all individual modeling groups is key to resolving differences in interpretation of implementation (scenarios and parameters) because the models generally work differently. Such a pipeline is important for improving the efficiency with which ensemble results are produced. Once ensemble results are confirmed, straightforward visualizations of results can be produced for decision-makers that illustrate the benefit of reducing modeling uncertainty given outbreak measures of interest (such as Tables 2 and 3) and that illustrate the

relative benefit of different control actions while integrating across models and incorporating our uncertainty in predictions (such as Figure 2). Our experience has been that both modeling groups and data owners are fundamentally interested in collaboration and quickly see the benefits of EM and SDM approaches, but patience and persistence are needed to successfully develop the type of consortium needed to implement this framework.

Conclusions

Because an integrated EM and SDM framework will evaluate the outcomes of all models in an ensemble across multiple objectives, they are useful to highlight control actions that are robust to existing model uncertainty, identify the key differences among models in the ensemble that must be clarified to resolve uncertainty in the best action, and illustrate trade-offs among the objectives of management. Although we were motivated here by our experience with FMD models, the proposed framework is broadly applicable to most, if not all, transboundary animal diseases. Full development of this framework will take time, but it is a good investment because of the role of models in policy and the complexity of integrating outputs from multiple models. Clearly, there is a need to more strongly engage policy makers in development and use of more science-based processes to integrate model recommendations both to inform policy and to overcome constraints such as data collection and data sharing. Although many challenges exist to the development of ensemble approaches for models of livestock and other diseases, their successful application in weather forecasting and other predictive sciences provide strong evidence for the importance of pursuing similar approaches in disease modelling.

Acknowledgements

Funding provided by the Research and Policy for Infectious Disease Dynamics (RAPIDD) Program, Science and Technology Directorate, US Department of Homeland Security, and Fogarty International Center, National Institutes of Health through interagency agreement #HSHQDC-09-X-00135. We especially thank the AusSpread, CVI, Exodis FMD, InterSpread Plus, NAADSM and Warwick modelling teams for providing model outputs for our analyses and Kelly A. Patyk for her comments on the manuscript. MF and MT are funded by a grant from the Ecology and Evolution of Infectious Disease program of the NSF/NIH (award number 1 R01 GM105247-01).

References

- Anonymous, 2014. Terrestrial Animal Health Code 23rd edition. Office International des Épizooties, Paris.
- Araujo, M.B., New, M., 2007. Ensemble forecasting of species distributions. *Trends Ecol. Evol.* 22, 42-47.
- Argyris, C., Schön, D.A., 1978. Organizational Learning: a Theory of Action Learning. Addison-Wesley, Reading, MA.
- Backer, J.A., Hagenaars, T.J., Nodelijk, G., van Roermund, H.J.W., 2012. Vaccination against foot and mouth disease I: epidemiological consequences. *Prev. Vet. Med.* 107, 27–40.
- Barbet-Massin, M., Walther, B.A., Thuiller, W., Rahbek, C., Jiguet, F., 2009. Potential impacts of climate change on the winter distribution of Afro-Palaeartic migrant passerines. *Biol. Lett.* 5, 248-251.

480 Beckett, S.D., Garner, M.G., 2007. Simulating disease spread within a geographic information
 481 system environment. *Veterinaria Italiana*. 43, 595-604.

482 Benestad, R.E., 2004. Tentative probabilistic temperature scenarios for northern Europe. *Tellus*
 483 Series a-Dynamic Meteorology and Oceanography. 56, 89-101.

484 Catelaube, P., Terres, J.M., 2005. Seasonal weather forecasts for crop yield modelling in Europe.
 485 *Tellus* 57, 476-487.

486 Chandler, R.E., 2013. Exploiting strength, discounting weakness: combining information from
 487 multiple climate simulators. *Philos. Trans. A Math. Phys. Eng. Sci.* 371 20120287, doi:
 488 10.1098/rsta.2012.0287.

489 Clemen, R., 1997. *Making Hard Decisions*. Duxbury Press, Pacific Grove, California, USA.

490 Cloke, H.L., Pappenberger, F., 2009. Ensemble flood forecasting: A review. *J. Hydrol.* 375, 613–
 491 626.

492 Coetsee, B.W.T., Robertson, M.P., Erasmus, B.F.N., van Rensburg, B.J., Thuiller, W., 2009.
 493 Ensemble models predict Important Bird Areas in southern Africa will become less effective
 494 for conserving endemic birds under climate change. *Glob. Ecol. Biogeogr.* 18, 701-710.

495 DEFRA, 2005. Cost-benefit analysis of foot and mouth disease control. *Risk Solut.* D 5100/R3
 496 D5100/R3, 109.

497 Doucet, A., de Freitas, N., Gordon, N., Eds., 2001. *Sequential Monte Carlo Methods in Practice*.
 498 Springer-Verlag, New York.

499 Dubé, C., Garner, M.G., Teachman, M.E., Whilesmith, J.W., Griffing, J., van Haldere, A.,
 500 Stevenson, M.A., Sanson, R.L., Harvey, N., 2006. The Animal Health Quadrilateral Epiteam
 501 – International collaboration on Foot and Mouth Disease simulation modelling for emergency

502 preparedness. In Proc. 11th Symposium of the International Society for Veterinary
 503 Epidemiology and Economics (ISVEE), 6-11 August 2006 Cairns. ISVEE 11, 339-343.
 504 Dubé, C., Stevenson, M.A., Garner, M.G., Sanson, R.L., Corso, B.A., Harvey, N., Griffin, J.,
 505 Wilesmith, J.W., Estrada, C., 2007. A comparison of predictions made by three simulation
 506 models of foot and mouth disease. *N. Z. Vet. J.* 55, 280-8.
 507 Ferguson, N.M., Donnelly, C.A., Anderson, R.M., 2001. The foot and mouth epidemic in Great
 508 Britain: Pattern of spread and impact of interventions. *Science* 292, 1155-1160.
 509 Garner, M.G., Beckett, S.D., 2005. Modelling the spread of foot and mouth disease in Australia.
 510 *Aust. Vet. J.* 83, 758-766.
 511 Garner, M.G., Hamilton, S.A., 2011. Principles of epidemiological modelling. *Rev. Sci. Tech.*
 512 30, 407-416.
 513 Gerber, L.R., Wielgus, J., Sala, E., 2007. A decision framework for the adaptive management of
 514 an exploited species with implications for marine reserves. *Conserv. Biol.* 21, 1594-1602.
 515 Gloster, J., Jones, A., Redington, A., Burgin, L., Sorensen, J.H., Turner, R., Dillon, M.,
 516 Hullinger, P., Simpson, M., Astrup, P., et. al., 2010. Airborne spread of foot and mouth
 517 disease - Model intercomparison. *Vet. J.* 183, 278-286.
 518 Gneiting, T., Raftery, A.E., 2005. Atmospheric science: Weather forecasting with ensemble
 519 methods. *Science* 310, 248-249.
 520 Green, L.E., Medley, G.F., 2002. Mathematical modelling of the foot and mouth disease
 521 epidemic of 2001: Strengths and weaknesses. *Res. Vet. Sci.* 73, 201-205.
 522 Guis, H., Caminade, C., Calvete, C., Morse, A.P., Tran, A., Baylis, M., 2012. Modelling the
 523 effects of past and future climate on the risk of bluetongue emergence in Europe. *J. R.*
 524 *Soc. Interface* 9, 339-350.

525 Harvey, N., Reeves, A., Schoenbaum, M.A., Zagmutt-Vergara, F.J., Dubé, C., Hill, A.E.,
 526 Corso, B.A., McNab, W.B., Cartwright, C.I., Salman, M.D., 2007. The North American
 527 Animal Disease Spread Model: a simulation model to assist decision making in evaluating
 528 animal disease incursions. *Prev. Vet. Med.* 82, 176-197.

529 Hayama, Y., Muroga, N., Nishida, T., Kobayashi, S., Tsutsui, T., 2012. Risk factors for local
 530 spread of foot and mouth disease, 2010 epidemic in Japan. *Res. Vet. Sci.* 93, 631-635.

531 Hollings, E.F., 1978. Resolving conflicting uses of oceans is most important. *Sea Technology* 19,
 532 20.

533 Kao, R.R., 2002. The role of mathematical modelling in the control of the 2001 FMD epidemic
 534 in the UK. *Trends in Microbiol.* 10, 279-286.

535 Keeling, M.J., 2005. Models of foot and mouth disease. *Proc. Biol. Sci.* 272, 1195-1202.

536 Keeling, M.J., Woolhouse, M.E., May, R.M., Davies, G., Grenfell, B.T., 2003. Modelling
 537 vaccination strategies against foot and mouth disease. *Nature* 421, 136-142.

538 Keeling, M.J., Woolhouse, M.E., Shaw, D.J., Matthews, L., Chase-Topping, M., Haydon, D.T.,
 539 Cornell, S.J., Kappey, J., Wilesmith, J., Grenfell, B.T., 2001. Dynamics of the 2001 UK foot
 540 and mouth epidemic: Stochastic dispersal in a heterogeneous landscape. *Science* 294, 813-
 541 817.

542 Keith, D.A., Martin, T.G., McDonald-Madden, E., Walters, C., 2011. Uncertainty and adaptive
 543 management for biodiversity conservation. *Biol. Conserv.* 144, 1175-1178.

544 Kraemer, M.U.G., Hay, S.I., Pigott, D.M., Smith, D.L., Wint, G.R.W., Golding, N., 2016.
 545 Progress and challenges in infectious disease cartography. *Trends in Parasitol.* 32: 19-29.

546 Kuhnert, P.M., Martin, T.G., Griffiths, S.P., 2010. A guide to eliciting and using expert
 547 knowledge in Bayesian ecological models. *Ecol. Lett.* 13, 900-914.

548 Lee, K.N., 1993. *Compass and Gyroscope: Integrating Science and Politics for the Environment*.
 549 Island Press, Washington, DC.

550 Lessler, J., Azman, A.S., Grabowski, M.K., Salje, H., Rodriguez-Barraquer, I., 2016. Trends in
 551 the mechanistic and dynamic modeling of infectious diseases. *Current Epidemiology Reports*
 552 3: 212-222.

553 Lindström, T., Tildesley, M., Webb, C., 2015. A Bayesian Ensemble Approach for
 554 Epidemiological Projections. *PLoS Comput. Biol.* 11, e1004187.

555 Maiorano, L., Falcucci, A., Zimmermann, N.E., Psomas, A., Pottier, J., Baisero, D., Rondinini,
 556 C., Guisan, A., Boitani, L., 2011. The future of terrestrial mammals in the Mediterranean
 557 basin under climate change. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 366, 2681-2692.

558 Mangiameli, P., West, D., Rampal, R., 2004. Model selection for medical diagnosis decision
 559 support systems. *Decis. Support Syst.* 36, 247-259.

560 Mansley, L.M., Donaldson, A.I., Thrusfield, M.V., Honhold, N., 2011. Destructive tension:
 561 mathematics versus experience--the progress and control of the 2001 foot and mouth disease
 562 epidemic in Great Britain. *Rev. Sci. Tech.* 30, 483-498.

563 McDonald-Madden, E., Probert, W.J.M., Hauser, C.E., Runge, M.C., Possingham, H.P., Jones,
 564 M.E., Moore, J.L., Rout, T.M., Vesk, P.A., Wintle, B.A., 2010. Active adaptive conservation
 565 of threatened species in the face of uncertainty. *Ecol. Appl.* 20, 1476-1489.

566 Morris, R.S., Wilesmith, J.W., Stern, M.W., Sanson, R.L., Stevenson, M.A., 2001. Predictive
 567 spatial modelling of alternative control strategies for the foot and mouth disease epidemic in
 568 Great Britain, 2001. *Vet. Rec.* 149, 137-144.

569 Murray, C.J., Ortblad, K.F., Guinovart, C., Lim, S.S., Wolock, T.M., Roberts, D.A.,
 570 Dansereau, E.A., Graetz, N., Barber, R.M., Brown, J.C., et al., 2014. Global, regional, and

571 national incidence and mortality for HIV, tuberculosis, and malaria during 1990-2013: A
 572 systematic analysis for the Global Burden of Disease Study 2013. *Lancet* 384, 1005-70.
 573 Murray, C.J., Rosenfeld, L.C., Lim, S.S., Andrews, K.G., Foreman, K.J., Haring, D., Fullman,
 574 N., Naghavi, M., Lozano, R., Lopez, A.D., 2012. Global malaria mortality between 1980 and
 575 2010: a systematic analysis. *Lancet* 379, 413–431.
 576 Niu, S., Luo, Y., Dietze, M.C., Keenan, T.F., Shi, Z., Li, J., Chapin, F.S. III, 2014. The role of
 577 data assimilation in predictive ecology. *Ecosphere* 5, article 65.
 578 Orsolini, Y.J., Doblas-Reyes, F.J., 2003. Ozone signatures of climate patterns over the Euro-
 579 Atlantic sector in the spring. *Q.J.R. Meteorol. Soc.* 129, 3251-3263.
 580 Paarlberg, P.L., Hillberg-Seitzinger, A., Lee, J.G., Mathews, K.M. Jr., 2008. Economic Impacts
 581 of Foreign Animal Disease. ERR-57. U.S. Dept. of Agriculture, Econ. Res. Serv.
 582 Palmer, T.N., Doblas-Reyes, F.J., Hagedorn, R., Alessandri, A., Gualdi, S., Andersen, U.,
 583 Feddersen, H., Cantelaube, P., Terres, J.M., Davey, M., et al., 2004. Development of a
 584 European multimodel ensemble system for seasonal-to-interannual prediction (DEMETER).
 585 *Bulletin of the American Meteorological Society.* 85, 853-872.
 586 Parma, A., 1999. What can adaptive management do for our fish, forests, food, and biodiversity?
 587 *Integr. Biol.* 1, 16-26.
 588 Probert, W.J.M., K. Shea, C.J. Fonnesbeck, M.C. Runge, T.E. Carpenter, S. ü, M.G. Garner, N.
 589 Harvey, M.A. Stevenson, C.T. Webb, M. Werkman, M.J. Tildesley, M.J. Ferrari, 2016.
 590 Decision-making for foot-and-mouth disease control: Objectives matter. *Epidemics* 15, 10-19.
 591 Roche, S.E., Garner, M.G., Wicks, R.M., East, I.J., de Witte, K., 2014. How do resources
 592 influence control measures during a simulated outbreak of foot and mouth disease in
 593 Australia? *Prev. Vet. Med.* 113, 436-446.

594 Roche, S.E., Garner, M.G., Sanson, R.L., Cook, C., Birch, C., Backer, J.A., Dube, C., Patyk,
 595 K.A., Stevenson, M.A., Yu, Z.D., Rawdon, T.G., Gauntlett, F. 2015. Evaluating vaccination
 596 strategies to control foot-and-mouth disease: a model comparison study. *Epidemiol. Infect.*
 597 143, 1256-1275.

598 Sanders, F., 1963. On subjective probability forecasting. *Journal of Applied Meteorology* 2, 191-
 599 201.

600 Sanson, R.L., Harvey, N., Garner, M.G., Stevenson, M.A., Davies, T.M., Hazelton, M.L.,
 601 O'Connor, J., Dubé, C., Forde-Folle, K., Owen, K., 2011. Foot and mouth disease model
 602 verification and 'relative validation' through a formal model comparison. *Rev. Sci. Tech.* 30,
 603 527-540.

604 Schoenbaum, M.A., Disney, W.T., 2003. Modelling alternative mitigation strategies for a
 605 hypothetical outbreak of foot and mouth disease in the United States. *Prev. Vet. Med.* 58, 25-
 606 52.

607 Shaman, J., Karspeck, A., 2012. Forecasting seasonal outbreaks of influenza. *Proc. Natl. Acad.*
 608 *Sci. USA.* 109. 20425–20430.

609 Shea, K., Management, N.W.G.P., 1998. Management of populations in conservation, harvesting
 610 and control. *Trends Ecol. Evol.* 13, 371-375.

611 Shea, K., Possingham, H.P., Murdoch, W.W., Roush, R., 2002. Active adaptive management in
 612 insect pest and weed control. Intervention with a plan for learning. *Ecol. Appl.* 12, 927-936.

613 Shea, K., Tildesley, M.J., Runge, M.C., Fonnesbeck, C.J., Ferrari, M.J., 2014. Adaptive
 614 management and the Value of Information: Learning via intervention in epidemiology. *PLoS*
 615 *Biol.* 12, e1001970.

616 Smith, T., Ross, A., Maire, N., Chitnis, N., Studer, A., Hardy, D., Brooks, A., Penny, M.,
 617 Tanner, M., 2012. Ensemble modeling of the likely public health impact of a pre-erythrocytic
 618 malaria vaccine. *PLoS Med.* 9, e1001157.

619 Smith, R.L., Tebaldi, C., Nychka, D., Mearns, L.O., 2009. Bayesian modeling of uncertainty in
 620 ensembles of climate models. *J. Am. Stat. Assoc.* 104, 97–116.

621 Stevenson, M.A., Sanson, R.L., Stern, M.W., O'Leary, B.D., Sujau, M., Moles-Benfell, N.,
 622 Morris, R.S., 2013. InterSpread Plus: a spatial and stochastic simulation model of disease in
 623 animal populations. *Prev. Vet. Med.* 109, 10-24.

624 Suen, S.C., Bendavid, E., Glodhaber-Fiebert, J.D., 2014. Disease control implications of India's
 625 changing multi-drug resistant tuberculosis epidemic. *PLoS One.* 9, e89822.

626 Taylor, N., 2003. Review of the use of models in informing disease control policy development
 627 and adjustment: A report for DEFRA.

628 Tebaldi, C., Knutti, R., 2007. The use of the multi-model ensemble in probabilistic climate
 629 projections. *Philos. Trans. A Math. Phys. Eng. Sci.* 365, 2053-2075.

630 Tebaldi, C., Smith, R.L., Nychka, D., Mearns, L.O., 2005. Quantifying uncertainty in projections
 631 of regional climate change : A Bayesian approach to the analysis of multimodel ensembles. *J.*
 632 *Clim.* 18, 1524–1540.

633 Thomson, M.C., Doblas-Reyes, F.J., Mason, S.J., Hagedorn, R., Connor, S.J., Phindela, T.,
 634 Morse, A.P., Palmer, T.N., 2006. Malaria early warnings based on seasonal climate forecasts
 635 from multi-model ensembles. *Nature* 439, 576-579.

636 Thuiller, W., Lafourcade, B., Engler, R., Araujo, M.B., 2009. BIOMOD - a platform for
 637 ensemble forecasting of species distributions. *Ecography.* 32, 369-373.

638 Tildesley, M.J., Deardon, R., Savill, N.J., Bessell, P.R., Brooks, S.P., Woolhouse, M.E.J.,
 639 Grenfell, B.T., Keeling, M.J., 2008. Accuracy of models for the 2001 UK foot and mouth
 640 epidemic. *Proc. Biol. Sci.* 275, 1459-1468.

641 Tildesley, M.J., Savill, N.J., Shaw, D.J., Deardon, R., Brooks, S.P., Woolhouse, M.E.J.,
 642 Grenfell, B.T., Keeling, M.J., 2006. Optimal reactive vaccination strategies for a foot and
 643 mouth outbreak in the UK. *Nature* 440, 83-86.

644 Velázquez, J.A., Anctil, F., Perrin, C., 2010. Performance and reliability of multimodel
 645 hydrological ensemble simulations based on seventeen lumped models and a thousand
 646 catchments. *Hydrol. Earth Syst. Sci.* 14, 2303–2317.

647 Walters, C.J., 1986. *Adaptive Management of Renewable Resources*. Blackburn Press, Caldwell,
 648 New Jersey.

649 Ward, M.P., Highfield, L., Carpenter, T.E., Garner, M.G., Beckett, S.D., Laffan, S.W., 2007.
 650 Multi-model investigation of foot and mouth disease spread in Texas. *Prev. Vet. Med.* 81,
 651 221-222.

652 West, D., Mangiameli, P., Rampal, R., West, V., 2005. Ensemble strategies for a medical
 653 diagnostic decision support system: A breast cancer diagnosis application. *Eur. J. Oper. Res.*
 654 162, 532-551.

655 Wilder-Smith, A., Macary, P., 2014. Dengue: challenges for policy makers and vaccine
 656 developers. *Curr. Infect. Dis. Rep.* 16, 404.

657 Willeberg, P., Grubbe, T., Weber, S., Forde-Folle, K., Dubé, C., 2011. The World Organization
 658 for Animal Health and epidemiological modelling: background and objectives. *Rev. Sci.*
 659 *Tech.* 30, 391-405.

660 Williams, B.K., 2011. Adaptive management of natural resources--framework and issues. J.
 661 Environ. Manage. 92, 1346-1353.
 662 Williams, B.K., Eaton, M.J., Breininger, D.R., 2011. Adaptive resource management and the
 663 value of information. Ecol. Modell. 222, 3429-3436.
 664 Williams, B., Szaro, R., Shapiro, C., 2007. Adaptive management: the US Department of the
 665 Interior technical guide. US Department of Interior, Adaptive Management Working Group,
 666 editor. Washington, DC. pp. 72.
 667 Woolhouse, M.E.J., 2003. Foot and mouth disease in the UK: What should we do next time?
 668 Symp. Ser. Soc. Appl. Microbiol. 32, 126S-130S.
 669 Yamana, T.K., Kandula, S., Shaman, J., 2016. Superensemble forecasts of dengue outbreaks. J.
 670 R. Soc. Interface 13: 20160410.
 671 Yoon, H., Wee, S.H., Stevenson, M.A., O'Leary, B.D., Morris, R.S., Hwang, I.J., Park, C.K.,
 672 Stern, M.W., 2006. Simulation analyses to evaluate alternative control strategies for the 2002
 673 foot and mouth disease outbreak in the Republic of Korea. Prev. Vet. Med. 74, 212-225.

Appendix A - Methods for Case 1: Structured decision-making

For each of the 15 combinations of five control actions and three models (AusSpread, NAADSM, and the Warwick model), we generated 100 stochastic simulations of an FMD outbreak on a simulated landscape of 8000 farms. Farm sizes, composition (proportions sheep and cattle), and spatial distribution were chosen to be consistent with the Cumbria region of the UK. We chose the Cumbria region because of its relevance for the 2001 UK FMD outbreak, and because the models used in this example were already parameterized for an FMD outbreak in this region. During the UK 2001 outbreak, Cumbria was severely affected, with between 20 and 30 farms reporting infection per day at the peak of the outbreak and animals on up to 150 farms being pre-emptively culled in an attempt to control the outbreak. This resulted in a maximum of 48,000 animals being culled per day in Cumbria alone. Vaccination was not used in 2001 for a number of reasons, not least of which was that there was insufficient capacity at the time to carry out a sustained vaccination campaign (Andersen 2002). Since 2001, vaccination has been considered as part of the UK FMD contingency plan, with DEFRA estimating that at most 35,000 animals could be vaccinated per day nationwide during a future FMD epidemic (Tildesley et al. 2006). In this paper we are considering a localised outbreak in Cumbria from a single source and with this in mind we assume a conservative daily culling capacity of 50 farms per day and a maximum vaccination capacity of 10,000 animals per day. Our objective in this section of the paper is to explore the effectiveness of structured decision making in determining the effectiveness of control, and it would be naïve to assume that the optimal strategy will be consistent as capacities are increased.

For all simulations we assumed an initial period of undetected spread for 10 days prior to the first detected case. Parameterizations for NAADSM and AusSpread were based on those

described in Sanson et al. (2011). The parameterization used in the Warwick model was as in (Tildesley et al., 2008). The reimbursement costs to farmers were calculated as 1000£ per cattle and 100£ per sheep and are based upon estimates of market prices of cattle and sheep in the UK during the 2001 outbreak.

Appendix B - Methods for Case 2: Determining ensemble weights

Application of Bayesian Reliable Ensemble Average Method to Epidemiology

We here describe the BREA method used in Case study 2. For a fuller exposition on BREA methods in epidemiology including both bias and convergence criteria, we refer readers to Lindström et al. (2015).

One of the key aspects of the BREA method is that weights, expressed as a precision parameter λ_i , are estimated jointly with the parameters of interest. In the original climate-change application of the BREA method (Tebaldi et al., 2005), the main quantity of interest was the estimated current and future mean temperature, denoted μ and $\hat{\mu}$ respectively. The relationship between these quantities (included in the analysis as random variables) and simulated current and future mean temperatures (denoted X_i and Y_i , respectively) for each model i was given by

$$\begin{aligned} X_i &\sim \text{Normal}(\mu, \lambda_i^{-1}) \\ Y_i &\sim \text{Normal}(\nu + \beta(X_i - \mu), (\theta\lambda_i)^{-1}) \end{aligned} \tag{0.1}$$

The parameter β is included to allow for correlation between current and future temperature estimates, so that, for example, a model that under-predicts current mean temperatures might also do so for future mean temperatures. Further, θ is included to allow for different levels of

discrepancy between projections of current and future temperatures, e.g. model simulation outputs may be more similar for current than for future temperature projections.

The BREA climate-change example is analogous to the epidemiological problem where, instead of current and future mean temperatures, we substituted an outbreak summary statistic (e.g., number of culled animals, number of vaccine doses administered, outbreak duration) under two different control actions. For equal-weighting of models, we estimated a single precision parameter $\hat{\lambda}$, common for all models, i.e. $\lambda_1 = \lambda_2 = \dots \lambda_n = \hat{\lambda}$, and for weights based on the convergence criterion we estimated λ_i for each model i . For the latter we also implemented a hierarchical approach similar to Smith et al. (2009) with $\lambda_i \sim \text{Gamma}(k_\lambda, k_\lambda/m_\lambda)$ that estimates hyperparameters k_λ (shape) and m_λ (mean) of λ in the analysis (Lindström et al., 2015). This corresponds to the assumption that the models in the ensemble come from a population of possible models, and the outbreak quantities of interest for this population are estimated. This approach reduces the sensitivity to which models are included or excluded in the analysis (Smith et al., 2009). Defining the gamma distribution by m_λ allows us to specify a prior for a hyperparameter that corresponds to $\hat{\lambda}$ in the equal-weighting analysis.

The method proposed by Tebaldi et al. (2005) also includes observed mean temperature, X_0 , in the analysis as $X_0 \sim \text{Normal}(\mu, \lambda_0^{-1})$ where λ_0 is the precision of natural variability in temperature. In climate modelling, it is reasonable that λ_0 is known, and it might also be the case for some data-rich diseases that variability in outbreak size or duration is known. However, in other cases such as FMD, natural variability in outbreak summary statistics is unknown. Thus, we included $\lambda_0 \sim \text{Gamma}(a_\tau, b_\tau)$ as an estimated parameter for the natural variability in the outbreak summary statistic in the epidemiological application of BREA (Lindström et al., 2015).

The stochastic simulations used for projection provided a mean simulated summary statistic, but also a range of the summary statistic. In the absence of a sufficient number of observed outbreaks to quantify λ_0 , we estimated λ_0 based on variability in the simulated projections via the hierarchical parameters, a_τ, b_τ .

Because the BREA method is a Bayesian approach, priors need to be specified for all random variables. Where possible, we implement the same, vague priors as used by Tebaldi et al. (Tebaldi et al., 2005) and specified $P(\mu) = P(\nu) = P(\theta) \propto 1$ and $P(\beta) = \text{Gamma}(a_\beta, b_\beta)$, i.e. a gamma distribution with shape a_β and rate b_β , with $a_\beta = b_\beta = 0.001$. For the analysis of equal weights, we implemented the prior $P(\hat{\lambda}) = \text{Gamma}(a_{\hat{\lambda}}, b_{\hat{\lambda}})$, with $a_{\hat{\lambda}} = b_{\hat{\lambda}} = 0.001$. For the model with different weights, we implemented a hierarchical model, similar to Smith et al. (Smith et al., 2009), and specified $\lambda_i \sim \text{Gamma}(k_\lambda, k_\lambda/m_\lambda)$, i.e. a gamma distribution with shape k_λ and mean m_λ . By using this parameterization, we may express the prior on m_λ , which is the corresponding parameter to $\hat{\lambda}$ in the equal-weight analysis. Thus, by using $P(m_\lambda) = \text{Gamma}(a_m, b_m)$ for $a_m = b_m = 0.001$, we may ensure that potential differences observed between the two weighting schemes are not the result of different priors. We also specified $P(k_\lambda) = \text{Gamma}(a_k, b_k)$ for $a_k = b_k = 0.001$, thus allowing for a wide range of shapes of the hierarchical distribution.

Because duration is inherently positive, we specify our model on the log-scale to fit with the assumptions of Eq. 0.1. That is, X_i and Y_i are interpreted as the mean log-duration, and μ and ν are the corresponding ensemble quantities. In Figure 2, we present the marginal distribution of these quantities, i.e. integrating over all other parameters in Eq 0.1, including model weights λ_i . However, for transparency we transform all quantities and parameter estimates back to the

original scale (rather than the log-transformed duration) with days as unit. As such, our results are presented for the geometrical mean duration.

Simulations

Case study 2 focuses on a mock outbreak of FMD in a subpopulation of farms from the UK, consisting of the Midlands counties and Wales. AusSpread, the CVI model, Exodis FMD, InterSpread Plus, and NAADSM had already simulated outbreaks as part of the QUADS studies (Roche et al., 2014; 2015). We simulated the Warwick model for the same initial conditions, underlying demography, and control measures as the QUADS studies scenarios (as given in Roche et al., 2015). Table B1 summarizes the simulation data of the models used in the BREA analysis for Case Study 2.

Vaccinations included all species and were assumed to start 14 days after first detection. Simulations started after the silent-spread phase, thus excluding transmission via animal shipments, and all models, scenarios, and replicates were seeded with the same 20 infected farms, of which one was detected. Further details on the assumptions can be found in Roche et al. (2014; 2015).

Table 1. Summary of FMD model properties. All models are stochastic, spatially explicit, state–transition models. IP: infected premises, DC: dangerous contact, CP: contiguous premises.

Table 2. Mean predicted duration (days) of outbreak for each model and control action. Shading indicates the action resulting in the shortest predicted outbreak duration for each model.

Numbers in parentheses indicate the 10th and 90th quantiles of the distribution of outcomes. The “average” row gives results for an equally weighted mixture of the distributions resulting from each model.

Table 3. Model-averaged predicted cost for each objective (rows) and control action (columns).

Predicted costs are given in millions of pounds (£). Numbers in parentheses indicate the 10th and 90th quantiles of an equally-weighted mixture distribution of the outcomes of the three models. Shading indicates the action with lowest mean cost for each objective.

Table B1. Underlying data for Figure 2. Expected outbreak duration (log-transformed) under control actions with infectious premises culling (X) and with vaccination in addition (Y).

Figure 1. The distribution of predicted cattle culled for 100 realizations of each combination of model (rows) and control action (columns).

Figure 2. The expected predicted outbreak duration in days under control actions with infectious premises culling (A) and with vaccination in addition (B) and the difference from using vaccination (C). Coloured, dashed lines indicate the mean projection of each individual model,

24 consistently coloured across the three panels. The marginal posterior probabilities of the
25 ensemble analysis with equal weights (black lines) and convergence weighting (grey lines) are
26 indicated and were calculated as described in Appendix B.

27

Figure 1. The distribution of predicted cattle culled for 100 realizations of each combination of control action (rows) and model (columns).

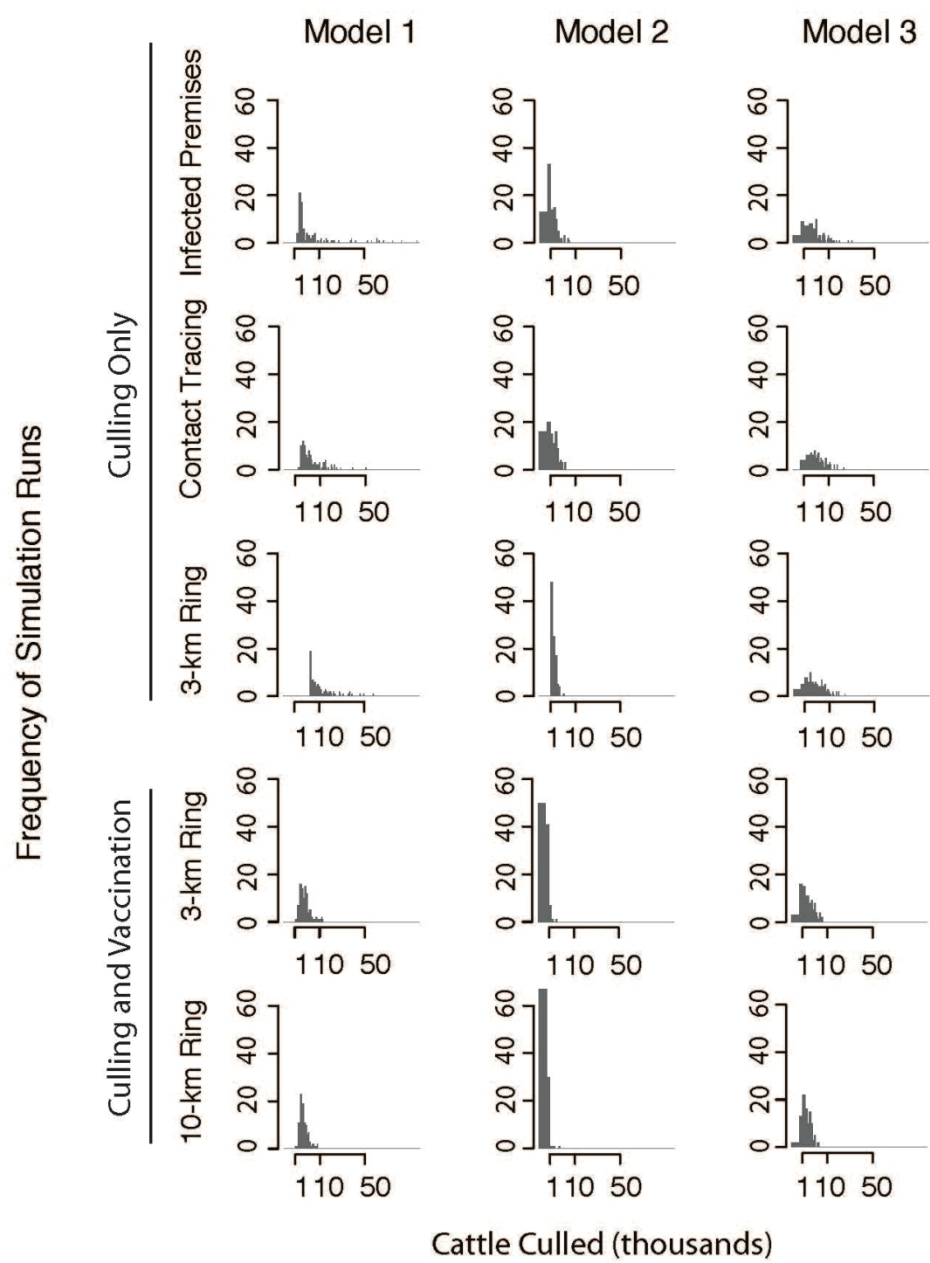


Figure 2. The expected predicted outbreak duration in days under control actions with infectious premises culling (A) and with vaccination in addition (B) and the difference from using vaccination (C). Coloured, dashed lines indicate the mean projection of each individual model, consistently coloured across the three panels. The marginal posterior probabilities of the ensemble analysis with equal weights (black lines) and convergence weighting (grey lines) are indicated and were calculated as described in Appendix B.

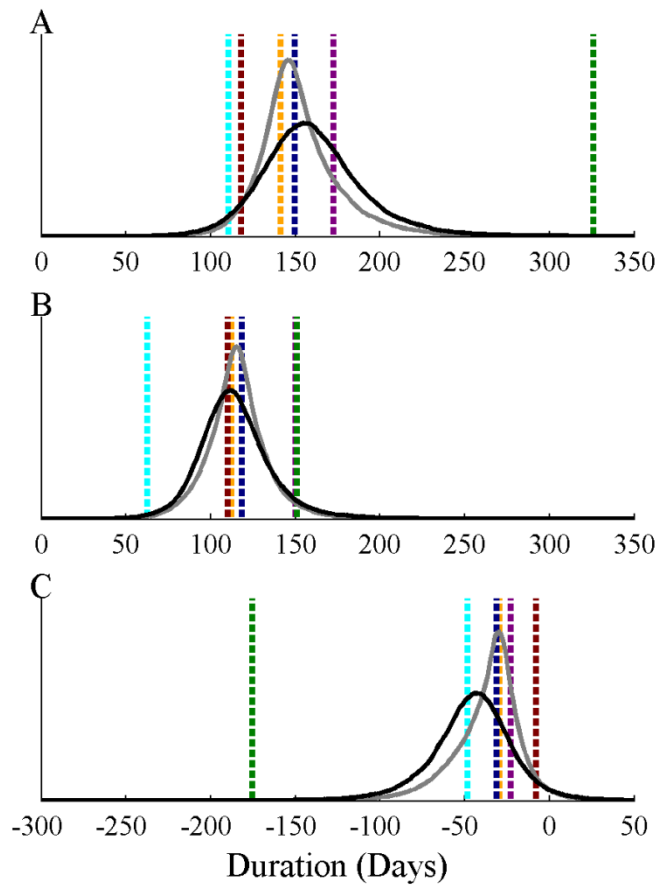


Table 1. Summary of FMD model properties. All models are stochastic, spatially explicit, state–transition models. IP: infected premises, DC: dangerous contact, CP: contiguous premises.

model	transmission via	control measures	references
AusSpread	Specific pathways	Quarantine, movement ban by zone or entire region, forward & backward tracing, IP, DC, and/or CP culls, vaccination, surveillance	Garner and Beckett, 2005; Beckett and Garner, 2007
CVI	Spatial kernel	Regulating transports, DC tracing, IP culls, ring culling, ring vaccination	Backer et al., 2012
Exodis-FMD	Mix of spatial kernel and specific pathways	Movement ban, protection & surveillance zones, culling of IP, DC, and/or contiguous, ring culling, welfare culling and vaccination, implemented by county.	DEFRA, 2005
InterSpread Plus	Specific pathways	Quarantine, movement ban by zone or entire region, forward & backward tracing, IP, DC and/or CP culls, vaccination, surveillance	Morris et al., 2001; Martinez-Lopez et al., 2009a; 2009b; Yoon et al., 2006; Stevenson et al., 2013

NAADSM	Specific pathways	Movement ban by entire region, forward tracing, IP, DC, and/or CP culls, vaccination, surveillance	Harvey et al., 2007
Warwick	Spatial kernel	Movement bans, IP, DC, and/or CP culls, vaccination	Keeling et al., 2001; Tildesley et al., 2006

Table 2. Mean predicted duration (days) of outbreak for each model and control action. Shading indicates the action resulting in the shortest predicted outbreak duration for each model.

Numbers in parentheses indicate the 10th and 90th quantiles of the distribution of outcomes. The “average” row gives results for an equally weighted mixture of the distributions resulting from each model.

	culling only			culling and vaccination	
	infected	contact	3-km ring	3-km	10-km
	premises ¹	tracing ²	culling ³	vaccination ⁴	vaccination ⁵
Mean predicted duration (days):					
Model 1	151 (39, 396)	98 (37, 182)	42 (23, 74)	69 (38, 101)	69 (34, 110)
Model 2	135 (59, 245)	137 (52, 243)	17 (11, 27)	116 (48, 213)	110 (45, 205)
Model 3	65 (27, 107)	42 (27, 56)	69 (29, 111)	43 (23, 64)	38 (24, 49)
average	117 (36, 222)	92 (33, 187)	43 (13, 93)	76 (30, 159)	72 (29, 128)

¹ culling of infected premises only

² culling of infected premises and those identified as dangerous contacts

³ culling in a 3-km ring around infected premises, including infected premises

⁴ vaccination in a 3-km ring around infected premises and culling of infected premises

⁵ vaccination in a 10-km ring around infected premises and culling of infected premises

Table 3. Model-averaged predicted cost for each objective (rows) and control action (columns).

Predicted costs are given in millions of pounds (£). Numbers in parentheses indicate the 10th and

90th quantiles of an equally weighted mixture distribution of the outcomes of the three models.

Shading indicates the action with lowest mean predicted cost for each objective.

objective	<u>culling only</u>			<u>culling and vaccination</u>	
	infected	contact	3-km ring	3-km	10-km
	premises ¹	tracing ²	culling ³	vaccination ⁴	vaccination ⁵
Predicted costs in millions of pounds (£)					
vaccinate-to-live	11.0 (2, 19)	8.8 (2, 18)	10.6 (3, 20)	5.1 (2, 9)	4.5 (2, 8)
vaccinate-to-die	11.0 (2, 19)	8.8 (2, 18)	10.6 (3, 20)	23.8 (7, 44)	90.3 (22, 156)

¹ culling of infected premises only

² culling of infected premises and those identified as dangerous contacts

³ culling in a 3-km ring around infected premises, including infected premises

⁴ vaccination in a 3-km ring around infected premises and culling of infected premises

⁵ vaccination in a 10-km ring around infected premises and culling of infected premises

Table B1. Underlying data for Figure 2. Expected outbreak duration (log-transformed) under control actions with infectious premises culling (X) and with vaccination in addition (Y).

Model	1	2	3	4	5	6
X	5.0097	5.7874	4.7045	4.9517	5.1512	4.7702
Y	4.7761	5.0168	4.3199	4.7196	5.0105	4.7035

Title: Ensemble Modelling and Structured Decision-making to Support Emergency Disease
Management

Colleen T. Webb^a, Matthew Ferrari^b, Tom Lindström^{a,c}, Tim Carpenter^d, Salome Dürr^e, Graeme
Garner^f, Chris Jewell^g, Mark Stevenson^h, Michael P. Wardⁱ, Marleen Werkman^j, Jantien Backer^j
and Michael Tildesley^k

^aDepartment of Biology, Colorado State University, Fort Collins, CO USA

^bCenter for Infectious Disease Dynamics, Pennsylvania State University, University Park, PA
USA

^cIFM, Theory and Modelling, Linköpings Universitet, Linköping, Sweden

^dEpiCentre, Massey University, Palmerston North, New Zealand

^eVeterinary Public Health Institute, Vetsuisse Faculty, University of Berne, Switzerland

^fAnimal Health Policy Branch, Department of Agriculture, Canberra, Australia

^gInstitute of Fundamental Sciences, Massey University, Palmerston North, New Zealand

^h Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Parkville,
Victoria 3010, Australia

ⁱFaculty of Veterinary Science, The University of Sydney, Camden, Australia

^jCentral Veterinary Institute part of Wageningen UR (CVI), Lelystad, the Netherlands

^kWarwick Infectious Disease Epidemiology Research (WIDER) group, School of Life Sciences
and Mathematics Institute, University of Warwick, Coventry, UK

Corresponding author: Colleen T. Webb; colleen.webb@colostate.edu